AD-A102 169 STANFORD UNIV CA DEPT OF STATISTICS F/G 12/1

A PARAMETRIC ANALYSIS OF STRUCTURED AND UNSTRUCTURED Q-SORT DAT--ETC(U)

JUN 81 W D HEAVLIN N00014-76-C-0475

UNCLASSIFIED TR-303 NL

1 OF 1
AD
A102169

END
DATE
FILMED
8 81
DTIC

AD A102169

A PARAMETRIC ANALYSIS OF STRUCTURED AND UNSTRUCTURED Q-SORT DATA

By

William D. Heavlin

Accession For
NTIS GRA&I ☒
DTIC TAB ☐
Unannounced ☐
Justification

By
Distribution/
Availability Codes

Dist	Avail and/or Special
A	

TECHNICAL REP[redacted]T NO. 303

Jun[redacted]81

Prepared Under Contract
N00014-76-C-0475 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

TABLE OF CONTENTS

Page

Page

Introduction.

The present work develops an analysis for a particular kind of data called the Q-sort. Most commonly occurring in connection with personality assessment, this data is typically generated as follows: A rater is presented with a deck of cards — called the Q-deck or Q-set — on which are written different descriptive statements. The rater is told to order the cards according to some criterion. As a very common example, the cards of the Q-deck might have different descriptions of personality and the rater would be asked to order them according to their similarity to the personality of a designated individual — the subject. Although occasionally the rater's task is to completely order the cards — called Q-items or simply items — this procedure becomes far too demanding as the number of items increases. In the latter instance, the rater is asked to classify each item according to its degree of concordance with the subject, ties permitted, thereby making the rater's task tractable. However, in order to enforce the similarity of this simplified task to the more difficult task of completely ordering the items, the so-called forced distribution is imposed. Under this restriction, the number of items that the rater may assign to any rank is fixed.

For example, the number of Q-items in the deck is often 100. Nine categories of similarity might be used, ranging from 1 — "most uncharacteristic," through 5 — "neither characteristic nor uncharacteristic", up to 9 — "most characteristic". The number permitted in each of the nine categories might then be 5, 8, 12, 16, 18, 16, 12, 8, 5, respectively. Thus, exactly five items would be forced to be rated "most uncharacteristic",

eight other items forced to be rated at the next most uncharacteristic level, and so on. The rankings of the deck as a whole is called a Q-sort.

The rationale for such a forced distribution is that, like the complete ordering of the items, the moments of the distribution of scores of the items of any Q-sort are fixed. Noting that, in particular, within each subject the mean and variance of the items' scores are fixed, Q-sort data is sometimes described as being "standardized within subjects" as opposed to being "standardized within variables", the consequence of imposing the more usual location and scale invariance on a set of multivariate data.

The feature of Q that the psychometric community considers distinguishing is usually described in terms of the matrix of data, X, whose rows represent the different subjects and whose columns represent the different scores of the items. Thus, X_{ji} is the score given to the i-th item in the forced distribution describing the j-th subject. The common and familiar practice is to standardize the matrix X by arguing that one's inferences ought not depend on the overall level of the item (or variable) i; that is, one ought be invariant to $\overline{X}_{\cdot i}$. Similarly, the second central moment of the i-th item is usually considered an invariant. If one denotes the standardized version of X by Z, note that the correlation matrix of the items corresponding to X is simply $R = \frac{1}{J} Z'Z$. One might then decompose this matrix R into factors that, appropriately rotated, would reveal groups of similar items.

In contrast, with Q-sort data the quantities $\overline{X}_{j\cdot}$ are all constant and equal to the mean of the forced distribution, as are the analogous

second moments. With this observation, one might as well standardize X such that $\overline{X}_{j.} = 0$ and $\frac{1}{I} \Sigma_i X_{ji}^2 = 1$. Calling this transformed matrix Y, by analogy to the correlation matrix of the items, one considers the matrix $Q = \frac{1}{I} YY'$, the "correlation matrix of subjects." The matrix Q may then be decomposed by factor analytic techniques to obtain factors or "clusters" of similar subjects. For this reason, Q methodology is sometimes considered as a competitor to cluster analyses, or, rather as a forerunner of historical interest. Q does not explicitly formulate this problem as a clustering problem; as such, this methodology is rarely used (Overall and Klett [1972]).

However, the factor analysis of subject-standardized Q data was not the only, nor even the primary, proposal made by the innovator of Q, William Stephenson. His more fundamental contribution was the methodology whereby the Q-deck was itself constructed. These Q-decks are called structured, and are, historically, the first kind of Q-sets employed.

The starting point of the structured Q-set is the psychological theory whose validity is being investigated. In the area of personality theory, the type psychologies furnish the simplest examples. In such schemes, Q-items are chosen to represent different types postulated by a particular theory. By such a deliberate procedure, a design matrix Q can be designated as corresponding to the structure of the Q-deck. Stephenson himself typically created multiway cross-factorial designs, taking pains to "balance" the structure by ensuring each cell in such a design had an equal number of representative items. The reader is referred to Kerlinger (1972), for a detailed description of Stephenson's structured Q-sort methodology.

At the same time as Stephenson was developing various aspects of his Q-technique, an alternative paradigm for questionnaire construction was becoming

widely accepted. This paradigm was based on the dual concepts of validity and reliability, each of which was in turn refined into secondary levels — construct validity, concurrent validity, interrater reliability, intrarater reliability and so on. These concepts as a whole were integrated by Cronbach et al (1973) into a theory of generalizability.

Stephenson's structured Q-set (Stephenson [1953]) failed to successfully compete with the requirements of generalizablity theory. The methodological problems regarding its validity and reliability (e.g. Sundland [1962]) were sufficient to greatly restrict its use. In fact, Block (1961) was able to substantially alter the scope of Q-studies by responding to these issues of validity and reliability; the result was his unstructured California Q-set. Only after Block's work did Q become identified exclusively as the kind of factor/cluster analysis described above; other innovations of Stephenson, especially his structuring of Q-sets, received less attention.

Although the present work ultimately develops recommendations for unstructured Q-sort data, its fundamental import is a parametric model for structured Q-sorts. Key to the development of this analysis is the derivation of a sampling (or probability) function. The sampling function that is derived describes the probability that a given individual will give any particular response (i.e. any particular ordering of Q-items). This object ties the Q-sort to other preference ordering and selection models. Because an analogous preference ordering problem was posed and then solved by Luce (1959), a brief description of some of Luce's results are presented prior to the main body of chapter I. Against this background

an axiom similar to Luce's choice axiom is postulated; from this axiom the functional form of the sampling function is derived.

In chapter II the sampling function is reparametrized to make the statistical model parsimonious. The essential problem posed by the structured Q-set — the relation of the item design matrix to the subject design matrix — is implicit in this reparametrization. In addition, various modifications to the sampling function are proposed to facilitate its computation. Each of these modifications transforms the sampling function into a kind of conditional sampling function.

On the basis of chapter II, conditional likelihood functions can be formed. In chapter III, these conditional likelihoods become the objective functions which, when maximized, furnish estimates of the parameters. The consistency and asymptotic normality of these estimates are immediate consequences of Andersen (1970).

Chapter IV illustrates the manner in which the results of the previous chapters help to solve the inferential problems of the structured Q-studies. Interestingly, while the evaluation of the significance of "nuisance" effects conforms to the framework of the generalized likelihood ratio tests, the central hypothesis of structured Q-studies, the retrospective validity of the Q-set, does not. A modification is proposed that enables the evaluation of this hypothesis.

Chapter V develops a latent factor model appropriate for the analysis of unstructured Q-sorts. This model compares to that for the structured Q-sort as the usual multivariate factor model (Anderson [1958], chapter 11) compares to the multivariate general linear model (Anderson [1958], chapter 8). Chapter VI presents an example that illustrates the kind of analysis

this work makes possible. In the conclusion, we return to assess the consequences of the results of the present work.

I. Derivation of the sampling function.

The present chapter develops a model that idealizes the process by which a single individual sorts the Q-set. This model describes the stochastic process of the sorting — but only in a sense. For unlike a modeling of the sorting process per se, this idealization does not depend on any initial conditions, e.g. the initial ordering of the Q-items; therefore it is considerably simpler.

I.A. Luce's theory of choice behavior.

The model of the sorting process, with its derivation, is in many ways parallel to that of Luce (1959), who idealized the process of choosing the single most preferred object from among N such objects. Indeed, by developing a model of the Q-sorting process a new perspective is gained on Luce's model; a perspective not found in the literature of mathematical psychology, including the latter-day review of Luce (1977). To facilitate a comparison, Luce's work is briefly reviewed in this section.

I.A.1. Notation for Luce's model.

Let $T = \{x,y,z,\ldots,t\}$ be the (finite) set of objects under consideration. T is referred to as the universe.

Suppose $A \subseteq S \subseteq T$. Let $P_S(A)$ denote the probability that the object chosen as most preferred is an element of A when the selection offered was all elements in S.

I.A.2. Luce's choice axiom and its consequences.

Luce's axiom consists of the following assertion:

$$P_T(A) = P_S(A)\, P_T(S) \quad \text{for all} \quad A \subseteq S \subseteq T\,, \tag{I.A.2(1)}$$

a statement very similar to that of conditional probability, were $P_S(A) = P(A|S)$. (Unlike the rules of conditional probability, the axiom I.A.2(1) applies only to nested sets, $A \subseteq S \subseteq T$.)

A derived but equivalent form of the axiom is

$$\frac{P_T\{x\}}{P_T\{y\}} = \frac{P_S\{x\}}{P_S\{y\}} \quad \text{for all} \quad S \quad \text{such that} \quad \{x,y\} \varepsilon S\,, \tag{I.A.2(2)}$$

subject to regularity conditions that prevent division by zero. As a direct consequence of I.A.2(2), one may conclude there exists a function $v\colon T \to (0,\infty)$, unique up to changes in scale, such that

$$\frac{P_T\{x\}}{P_T\{y\}} = \frac{P_S\{x\}}{P_S\{y\}} = \frac{v(x)}{v(y)}\,, \tag{I.A.2(3)}$$

whence

$$P_S\{x\} = \frac{v(x)}{\sum\limits_{u \varepsilon S} v(u)}\,.$$

I.A.3. Interpretation of Luce's axiom.

Luce's axiom, in the form of I.A.2(2), is sometimes described as expressing a notion of "independence of irrelevant alternatives," with the following meaning: Suppose in the course of selecting the most preferred object from the set S, the choice narrows to one between elements x and y. Then the final decision is made by considering

only the merits of objects x and y; the properties of all other elements in S are "irrelevant".

Analogies to Luce's choice axiom and its consequences will be made in the following.

I.B. The exchange axioms.

Like Luce's axiomatization, the following focuses on the behavior of a single individual. The behavior, however, will consist of the sorting of the items of the Q-set. For the moment, and for present convenience only, the task will be to <u>completely</u> order the Q-items.

I.B.1. Notation.

Let I be the number of items and let the items be indexed $1,2,\ldots,I$. Let π, a permutation, represent an ordering of these items. The k-th component of π, denoted $\pi(k)$, is considered the index of the item ranked $I-k+1$. Thus,

$$\begin{aligned}
\pi(1) &= \text{index of the item ranked lowest ,} \\
\pi(2) &= \text{index of the item ranked second lowest ,} \\
&\vdots \\
\pi(I) &= \text{index of the item ranked highest .}
\end{aligned}$$

$p(\pi)$ is the probability that π will occur; $p(\pi)$ is called the sampling function.

Let ρ be the space of all permutations, π. Let ℓ and m be ranks such that $\ell > m$. Define the operator $\tau(\cdot;\ell,m)\colon \rho \to \rho$ as the one-to-one onto map such that

$$\begin{aligned}
&\tau(\pi;\ell,m)\ (k) = \pi(k) \text{ for all } k \neq \ell \text{ and } k \neq m \\
&\tau(\pi;\ell,m)\ (\ell) = \pi(m), \text{ and} \\
&\tau(\pi;\ell,m)\ (m) = \pi(\ell) \ .
\end{aligned}$$

Thus, $\tau(\pi;\ell,m)$ represents an ordering identical to that of π save that the indices $\pi(m)$ and $\pi(\ell)$ have been exchanged. This corresponds to the interchange of the two items, those ranked in the ℓ-th and m-th positions, in the ordering of the Q-set which π denotes.

I.B.2. The first exchange axiom and its consequences.

The first exchange axiom asserts

$$\frac{p(\pi)}{p(\tau(\pi;\ell,m))} = h(\pi(\ell),\pi(m);\ell,m) \quad \begin{array}{l} \text{for all } \ell > m, \\ \text{for all } \pi, \end{array} \qquad \text{I.B.2(1)}$$

with the following interpretation: The two permutations, π and $\tau(\pi;\ell,m)$, differ only in their placement of the items $\pi(\ell)$ and $\pi(m)$. Thus, in choosing between these two permutations, the decision intuitively ought not be based upon the properties of the other items. The ranks of all these other I-2 items are the same for the two permutations being compared; for this reason they are "irrelevant." Parallel to Luce's choice axiom, the notion of "independence of irrelevant alternatives" is the critical justification here.

A consequence of I.B.2(1) is that there exist positive parameters $\{\rho(i)\}$ and a function $S: \{1,\ldots,I\} \to (-\infty,\infty)$ such that

$$\frac{p(\pi)}{p(\tau(\pi;\ell,m))} = \left[\frac{\rho(\pi(\ell))}{\rho(\pi(m))}\right]^{S(\ell)-S(m)} . \qquad \text{I.B.2(2)}$$

(The essential point in the derivation is the observation that the ratio

$$\frac{p(\pi_0)}{p(\pi_1)} \frac{p(\pi_1)}{p(\pi_2)}$$

does not depend on π_1.) This statement similar in form to that of I.A.2(3), where the $\{\rho(i)\}$ correspond to the $\{v(x)\}$ of Luce's model. These parameters $\{\rho(i)\}$ will be referred to as "propensities", $\rho(i)$ the propensity of item i to being ranked highly. The function $S(\cdot)$ acts as a scaling function that defines the "distance" between the various ranks. Considerable discussion will be devoted to this scaling function in the remainder of this chapter.

The following axiom gives some insight into the role of $S(\cdot)$:

Define $\tau(\cdot;\ell): P \to P$ by

$$\tau(\pi;\ell)\ (k) = \pi(k)\ , \quad \text{for } k \neq \ell,\ k \neq \ell+1$$

$$\tau(\pi;\ell)\ (\ell) = \pi(\ell+1)\ ,$$

$$\tau(\pi;\ell)\ (\ell+1) = \pi(\ell)\ .$$

Thus $\tau(\pi;\ell)$ represents the exchange of two adjacent Q-items, i.e. the ℓ-th and $\ell+1$-st items.

The second exchange axiom, which is a specialization of the first, asserts that

$$\frac{p(\pi)}{p(\tau(\pi;\ell))} = h(\pi(\ell),\pi(\ell+1))\ , \qquad \text{I.B.2(3)}$$

which, like the first exchange axiom, asserts a notion of independence of irrelevant alternatives. But the consequences of I.B.2(3) are more severe; I.B.2(3) postulates there exist $\{\rho(i)\}$ such that

$$\frac{p(\pi)}{p(\tau(\pi;\ell))} = \frac{\rho(\pi(\ell+1))}{\rho(\pi(\ell))}\ , \qquad \text{I.B.2(4)}$$

a statement that is remarkably similar to I.A.2(3) and that constitutes a simplification of I.B.2(2). I.B.2(4) has $S(\ell+1) - S(\ell) = c_1$ and hence $S(\ell) = c_1(\ell-c_2)$, c_1,c_2 arbitrary constants. The role of $S(\cdot)$ then is that of a scaling function, as was mentioned above, measuring a sort of interval of discrimination between the ranks ranging from 1 to I. The strength, or severity, depending on one's point of view, of the second exchange axiom is its assertion that all the intervals of discrimination between the ranks are of equal importance. This assumption is generally not appropriate and will be modified below.

I.C. Secondary axioms and properties of the model.

A consequence of the first exchange axiom is that the functional form of $p(\cdot)$ is determined up to the parameters $\{\rho(i)\}$ and the scaling function $S(\cdot)$. Thus,

$$p(\pi) = \frac{\prod_{k=1}^{I} \rho(\pi(k))^{S(k)}}{\Sigma_{\pi'} \prod_{k'=1}^{I} \rho(\pi'(k'))^{S(k')}} \qquad \text{I.C.(1)}$$

where $\Sigma_{\pi'}$ denotes summation over all permutations π'. The denominator simply ensures that the probabilities sum to unity.

I.C.1. The monotone axiom.

Note that changes in scale in the $\{\rho(i)\}$ are equivalent to changes in location for $S(\cdot)$ while changes in power in the $\rho\{(i)\}$ are equivalent to changes in scale for $S(\cdot)$. In this sense, then, $S(\cdot)$ is determined up to affine transformations.

A natural regularity condition to postulate of $p(\pi)$ is for the ratio

$$\frac{p(\pi)}{p(\tau(\pi;\ell))} = \left(\frac{\rho(\pi(\ell+1))}{\rho(\pi(\ell))}\right)^{S(\ell+1)-S(\ell)} \qquad \text{I.C.1(1)}$$

to be increasing in $\rho(\pi(\ell+1))$. The rationale is the following: If $\rho(i)$ $(i=\pi(\ell+1))$ measures some propensity of the i-th

item to be ranked highly, then any increase in this propensity should be reflected in an increase in the likelihood of item i being ranked over any other item, in particular, over item $\pi(\ell)$.

If the ratio I.C.1(1) is to be increasing in $\rho(\pi(\ell+1))$, then $S(\ell+1) - S(\ell)$ must be positive; hence, $S(\cdot)$ must be monotone increasing. For this reason, I.C.1(1) is called the monotone axiom.

The monotone axiom enforces a property that parallels one in Luce's model, one directly deducible from Luce's choice axiom. This property, strong stochastic transitivity, is expressed in terms of the pairwise preference probabilities

$$P(x,y) = P_A\{x\} \quad \text{where} \quad A = \{x,y\} ,$$

recalling the notation of section I.A. The property of strong stochastic transitivity says that

$$\begin{array}{lll} \text{if} & P(x,y) \geq \frac{1}{2} & (y \text{ is not preferred to } x) \\ \text{and} & P(y,z) \geq \frac{1}{2} & (z \text{ is not preferred to } y) \end{array}$$

then $P(x,z) \geq \max\{P(x,y), P(y,z)\}$.

Objects such as these pairwise preferences, simple though they be, are not natural in the present context of the Q-sort problem.

However, the analogy is interesting. Note that the monotonicity of the ratio in I.C.1(1) is the same kind of property as that of strong stochastic transitivity in Luce's model — for essentially the same reason. For Luce's model, stochastic transitivity follows from the linear ordering of the objects that is induced by $v(\cdot)$. Similarly, for the Q-sorting model, a linear ordering is implicit in the scalar quantities $\{\rho(i)\}$. The monotone axiom ensures the linearity of this ordering.

I.C.2. The palindrome axiom.

Let us denote the ranking that is the reverse of that connoted by π as $\bar{\pi}$. Thus

$$\bar{\pi}(k) = \pi(I-k+1), \quad k = 1,\ldots,I .$$

The following axiom is sometimes reasonable:

$$\frac{p(\pi)}{p(\pi')} = \frac{p(\bar{\pi}')}{p(\bar{\pi})} \quad \text{for all } \pi , \pi' . \qquad \text{I.C.2(2)}$$

This axiom asserts the following kind of invariance: If the magnitude of the effects is reversed, and if the rankings that empirically measure these magnitudes is also reversed, no distortion in the structure of the probabilities would occur.

The condition that I.C.2(2) imposes upon the scaling function is that

$$S(k) = - S(I-k+1) , \quad k = 1,\ldots,I .$$

Such a property is commonly called "rank reversibility," although palindrome invariance has recently been suggested (McCullagh [1978]). For this reason, I.C.2(2) is called the palindrome axiom.

Luce's choice axiom has been shown inconsistent with this concept of palindrome invariance (Luce [1959] , Marley [1968]). Thus, the axiom I.C.2(2) represents a qualitative distinction between the Q-sorting model and Luce's model.

I.C.3. Axioms that determine the scaling function.

Until now, the sorting task has been assumed to be that of completely ordering the Q-items, a task very unrepresentative of standard Q-practice. This restriction was made for convenience only and this section will be devoted to relieving this restriction. Key to this discussion will be the choice of the scaling function.

The scaling function $S(\cdot)$ allows for adaptation of the model involving a complete ranking of the Q-items to the more common case involving sorting in accordance with a forced distribution. While the category sizes of the forced distribution can in principle be accounted for by a summation over all compatible rankings, such a summation is numerically complicated to implement. In addition, such a procedure fails to represent in the model the fact that the forced distribution is an <u>a priori</u>, designed feature of the sorting task.

As a simple alternative, one can represent the forced distribution by equating the values of the scaling function for ranks residing in the same category, thereby equating the ranks themselves. Were the sizes the categories 5,8,12,16,18,16,12,8, and 5, then S(1) through S(5) would have a common value, as would S(6) through S(13), S(14) through S(25),

and so on. While this procedure does not completely determine the form of $S(\cdot)$, it does greatly reduce its complexity.

By requiring the scaling function to be constant for ranks that share the same category in the forced distribution, another form of the exchange axiom is motivated, one as restrictive as the second exchange axiom but one that exploits the presence of the forced distribution. We make use of the following notation:

Let $\{C_k\}$ be mutually exclusive and exhaustive sets that partition the ranks 1 to I. Let the k-th of these sets, C_k, correspond to the k-th category of the forced distribution. For example, for the forced distribution with category sizes 5, 8, 12, 16, 18, 16, 12, 8, and 5, C_1 would be the set $\{1,2,\ldots,5\}$, C_2 the set $\{6,7,\ldots,13\}$, C_3 the set $\{14,15,\ldots,25\}$, and so on.

The notion of fixing the scaling function to a common value for the ranks in the same category of the forced distribution may be represented formally by the axiom

$$\frac{p(\pi)}{p(\tau(\pi;\ell,m))} = 1 \quad \text{for all } \ell, m \in C_k, \text{ for all } k, \text{ and for all } \pi . \qquad \text{I.C.3(1)}$$

The consequences of I.C.3(1) is that for all k

$$S(\ell) = S(m) \quad \text{whenever } \ell, m \in C_k .$$

We shall refer to the axiom I.C.3(1) as the first scaling axiom.

I.C.3(1) is the first axiom to be postulated that exploits the use of the forced distribution. All previous axioms, in particular

the first and second exchange axioms, were cast in the context of completely ordering the Q-set. Most notably, the second exchange axiom was able to largely determine the scaling function. The following assertion resembles the second exchange axiom but exploits the structure of the forced distribution:

Let $m \varepsilon C_k$ and $\ell \varepsilon C_{k+1}$. The second scaling axiom asserts that for all π,

$$\frac{p(\pi)}{p(\tau(\pi;\ell,m))} = h(\pi(\ell),\pi(m)) \quad . \qquad \text{I.C.3(2)}$$

(The distinction between I.C.3(2) and the second exchange axiom is that for I.C.3(2) ℓ and m are restricted to adjacent categories, while for the second exchange axiom, ℓ and m were adjacent ranks.) From I.C.3(2) it follows that if there is a k such that $m \varepsilon C_k$ and $\ell \varepsilon C_{k+1}$, then $S(\ell) - S(m)$ is a constant. As a consequence of I.C.3(2), the scaling function is essentially determined, that is, determined up to changes in location and scale.

A reasonable alternative to completely specifying the scale function is to estimate it statistically. This notion will be developed in the next chapter.

I.C.4. Context as defined by the Q-set.

One attraction of Q is the capacity to build and enforce a vocabulary; the introspection of the sorting process can be required to be done with reference to standard items. Stephenson

(1953) used this idea to build Q-sets specific to particular psychological theories, while Block (1961) used this idea in order to transcend the vocabularies of particular theories of personality — and especially to transcend idiosyncratic adaptations of these vocabularies. At the heart of this vocabulary-enforcing capacity is the idea of a "set", the global framework that personality and behavior inventories imply by asking the questions they do. Curiously, the formal model expressed by the sampling function in I.C(1) has an interesting property with regard to this concept of "set".

Nowhere in the derivation of the functional form of $p(\cdot)$ are the $\{\rho(i)\}$ defined <u>except in the context of all I items</u> simultaneously. Thus, were one to add a single item to the original Q-set or to delete one from the existing Q-set, nothing in the formal theory of the axioms presented above would allow one to infer that the properties of the original Q-set would be at all similar to the properties of the newly constituted Q-set. This formal non-correspondance of almost identical Q-sets sharply distinguishes the present model from that of Luce. In Luce's model, the choice axiom explicitly restricts the manner in which behavior, that is, the choice probabilities, can change in response to the context, i.e. the selection of choices available. Thus, Luce's model is a direct result of assuming stability as the context changes; the Q-sorting model, by contrast, makes no such assumption.

With this point in mind, the occasional practice of subsampling from a larger Q-set to ease the task of the rater (e.g. Jackson and Bidwell [1959]) requires — in the frame of this axiomatic development — an additional axiom to justify it. This axiom, which would assert

that $\rho(i)$ is the same for any Q-set in which i appears, could be appropriately called "context irrelevance." Further, if the items are chosen by explicitly sampling from a larger "population" of items, a similar axiom is required if generalizations to the item "population" as a whole are to be made.

The assumption of context irrelevance has practical consequences. One can easily imagine that some questionnaires achieve a certain "set" in their responders by asking certain questions in a certain order, a "set" that in turn can be reflected in their responses. Quite conceivably, to ask more or fewer questions, different questions in a different order, would achieve a different "set" and would result in different responses. Thus, although the issue is difficult to address experimentally, by no means is it insignificant.

II. The statistical model.

This chapter leaves behind the axiomatic development of the sampling function; the focus is exclusively upon the form of various (conditional) likelihood functions of the structured Q-sort. This focus is no doubt curious to some, for the likelihood functions furnish directly neither estimates nor inferential procedures. Only in chapter III will this deficiency be remedied; there estimates and inferences will be derived using the theory of maximum likelihood.

II.A. Model parametrization.

In this section the unconditional likelihood of the structured Q-sort will be derived.

II.A.1. Log-linear parameters and duality.

In chapter I the sampling function $p(\cdot)$ was developed for a single individual; it was parametrized by $\{\rho(i)\}$. To extend this sampling model to individuals $j = 1,\ldots,J$, the relevant parameters are $\{\rho(i,j)\}$. This is a very large number of parameters; the present task is to reduce the dimensionality of the parameters from the excessive number IJ to something smaller.

For the structured Q-sort, items are constructed to represent levels of various attributes. This structure is represented by an $I \times D$ design matrix, Q, whose rows, Q_i, are the indicators and level variables of the corresponding i-th Q-item. Q is called the <u>item design matrix</u>.

Let $\rho(i,j) = \exp\{Q_i\beta_j'\}$ where β_j is a $1 \times D$ vector consisting of the parameters of the j-th subject and act as the coefficients of the D variables that compose Q.

Finally, let $\beta_j' = \beta w_j'$, where w_j is a $1 \times K$ vector consisting of the covariates of the j-th subject, β the $D \times K$ matrix of unknown parameters. W, the $J \times K$ matrix whose rows are the w_j, is called the subject design matrix.

The interpretation of β is intriguing. As developed, the coordinates $\beta_j' = \beta w_j'$ *locate the j-th individual* in the (dual of the) design space of the Q-set spanned by the rows of the matrix Q. On the other hand, $\gamma_i = Q_i\beta$ represents the i-th Q-item in the (dual of the) coordinate space spanned by the rows of the matrix W. Thus, β represents each of these two linear spaces, the item design space and the subject design space, to one another by its respective rows and columns. This duality is reminiscent of the reciprocity principle upon which focused much debate about Q- versus R- factor analysis. (See Burt [1972] and Burt and Stephenson [1939].)

In this reparametrization, the sampling function of the j-th subject, $p_j(\cdot)$, becomes

$$p_j(\pi) = \frac{\prod_{k=1}^{I} \exp\{Q_{\pi(k)}\beta w_j'\}^{S(k)}}{\Sigma_{\pi'} \prod_{k'=1}^{I} \exp\{Q_{\pi'(k')}\beta w_j'\}^{S(k')}}, \qquad \text{II.A.1(1)}$$

$$= \frac{\exp\{q(\pi)\beta w_j'\}}{\Sigma_{\pi'} \exp\{q(\pi')\beta w_j'\}}$$

where $q(\pi) = \sum_{k=1}^{I} S(k)Q_{\pi(k)}$, a $1 \times D$ vector.

II.A.2. Refinements in the parametrization.

This section will develop two refinements of the log-linear parametrization that are directed at specialized concerns. Both are parsimonious and convenient to implement. In addition, each provides some insight into the workings of Q, and therefore provides useful criticism of the potential strengths and weaknesses of Q.

a. Parametrizing the scaling function.

In section I.C.1 the scaling function was observed to be determined up to affine transformations; in section I.C.2 it was postulated to be skew-symmetric. Because the scaling function can be interpreted as measuring the "subjective distances" between the sort's categories, some empirical validation of the scaling values used might be of interest. In particular, one might wish to determine if the discrimination between the extreme categories is greater or less than those between the neutral categories. Such an

issue can be formally considered by the following *association:* If the discrimination between the extreme categories is greater than that between the central categories, consider this well-expressed by postulating the scaling function to be convex above its median. If, conversely, the discrimination between the central categories are greater than those between the extreme categories, consider this well-expressed by postulating the scaling function to be concave above its median. The following parametrization is then motivated:

$$S_\alpha(k) = |S(k)|^\alpha \quad \mathrm{sgn}(S(k)) \ , \quad k = 1,2,\ldots,I$$

where $S(k)$ is any *a priori* skew-symmetric scaling function with zero median. Employing power transformations to skew-symmetric functions as surrogates to the wider class of functions that are monotone, skew-symmetric, and convex(concave)-above-the-median, this parametrization is especially parsimonious. As a side benefit, the monotone axiom can be evaluated. An estimate of α less than zero would indicate monotonicity was being strongly violated.

b. Modeling variations in raters.

In this section a particular kind of variation between raters is considered. This variation is due to the differing levels of effort that different raters are likely to make in producing their sorts.

For each rater r let $\alpha(r)$ be the "acuity" or effort parameter of that rater. Let $\rho(i,j,r)$ denote the propensity of item i being ranked highly when rater r rates subject j, while $\rho(i,j)$ becomes some "underlying" propensity of item i being ranked highly for subject j.

The modeling relation of the acuity of rater r is

$$\rho(i,j,r) = \rho(i,j)^{\alpha(r)} , \quad i = 1,\ldots,I .$$

For the familiar general linear model the analogue to this parametrization is that the residual variation is inhomogeneous and that this inhomogeneity can be attributed to the raters. More importantly, as the result of this analogy to residual variation, estimates of the $\alpha(r)$ can be used as indices of relative reliability that enable comparisons to a "standard" rater.

II.A.3. The unconditional or full *likelihood*.

Let the unconditional likelihood function be denoted by $e^{\mathcal{L}(\beta)}$. Then,

$$e^{\mathcal{L}(\beta)} = \prod_{j=1}^{J} p_j(\pi) = \prod_{j=1}^{J} \frac{\exp\{q(\pi_j)\beta w_j'\}}{\Sigma_{\pi'} \exp\{q(\pi')\beta w_j'\}} \qquad \text{II.A.3(1)}$$

where π_j now denotes the permutation observed as the response of the j-th subject. (Note that π_j, which connotes a full ranking of the I items of the Q-set, breaks ties arbitrarily. However, $q(\pi_j)$ is invariant to the manner in which the ties are broken.)

The rightmost expression in II.A.3(1) is uncomputable for even moderate I as it requires summations over all possible permutations of I objects. To avoid this problem, the remainder of this chapter considers various conditional likelihoods, each of which is potentially computable.

II.B. Model simplification.

Recall that the full likelihood was computationally intractable because the denominator of each $p_j(\pi)$ contained a very large number — I! — of terms. A natural simplification is to limit the number of terms in these denominators. The issue is then to determine the criterion by which these terms are chosen. Note that if one of these terms in the denominator of $p_j(\pi)$ is the numerator itself, two benefits accrue. First, $p_j(\pi_j)$ is bounded above by unity, thereby remaining a proper probability measure. Second, $p_j(\pi_j)$ can then be interpreted as a conditional probability.

The motivation here to employ conditional likelihoods is atypical. More commonly, conditional likelihoods are employed to eliminate so-called incidental parameters (Neyman and Scott [1948]) whose estimation would otherwise consume too many degrees of freedom. Alternatively, conditional likelihoods are sometimes explicitly induced by conditioning on those statistics that have no apparent relevance. (See, for example, Godambe [1980].) Although computational simplicity is sometimes listed as one of the virtues of conditional inference, it is usually subordinate to another criterion. Here, however, it must be pre-eminent.

In conceding this pre-eminence, the choice of the form of the conditional likelihood is largely undetermined. This section presents two kinds of conditional likelihoods. The first exploits the special structure that Stephenson built into his Q-sorts — a completely balanced, cross-factorial design. The second applies to more general Q-designs.

II.B.1. The balanced, cross-factorial designs.

Stephenson's Q-sets all had the following design: First, he would characterize a theory (of personality, of aesthetics, etc.) by F factors, the f-th of which would have L_f levels. Then, with an $L_1 \times L_2 \times \cdots \times L_F$ design in mind, he would develop a certain number of Q-items to be representative of the traits that characterize, according to the theory, each cell in this cross-factorial design. When each cell is represented by an equal number of Q-items the item design matrix is said to be balanced. Kerlinger (1972) gives a succinct but complete description of the ways in which Stephenson used these balanced, cross-factorial designs.

For clarity of presentation, let us consider $2 \times \cdots \times 2 = 2^F$ designs first. The convenience of doing this derives from the fact that in this case D = F. For $F = 3$, the Q matrix would have the typical rows

$$\begin{array}{rrr} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \\ -1 & -1 & -1 \end{array}$$

Thus each row represents a vertex of a cube whose center of mass is at the origin (0,0,0). Let the conditional likelihood for the balanced Q-sort, $e^{\mathcal{L}_{BC}(\beta)}$, then be

$$e^{\mathcal{L}_{BC}(\beta)} = \prod_{j=1}^{J} \left\{ \frac{\exp\{q(\pi_j)\beta w_j'\}}{\Sigma_S \exp\{q(\pi_j)S\beta w_j'\}} \right\} \qquad \text{II.B.1(1)}$$

where Σ_S denotes summation over all diagonal matrices S of order F with $+1$ and -1 being the only diagonal entries available. Recall that $q(\pi_j)$ are the marginal Q-scores of the D dimensions of the item design matrix. Thus,

$$q(\pi_j) = \sum_{k=1}^{I} S(k)\ Q_{\pi_j(k)} .$$

The above form corresponds to conditioning on

$$|q(\pi_j)| = (|q_1(\pi_j)|, |q_2(\pi_j)| ,\ldots, |q_F(\pi_j)|) .$$

Geometrically, this corresponds to conditioning on the event that $q(\pi_j)$ is one of the vertices of the right rectangular prism whose center of mass is at the origin and whose vertices are $(\pm |q_1(\pi_j)|, \pm |q_2(\pi_j)| ,\ldots, \pm |q_F(\pi_j)|)$. The rationale for conditioning as above is the following: (1) The vertices $\{q(\pi_j)S\}$ that compose the orbit of the conditioning have centroid at the origin and in that sense do not "bias" the likelihood in any particular direction. (2) The vertices $\{q(\pi_j)S\}$ span the design space whenever $q_f(\pi_j) \neq 0$ for all $f = 1,\ldots,F$; that is, these vertices span the design space whenever it is intuitively reasonable that it should, and otherwise they span the linear subspace of the design space that holds the information that $q(\pi_j)$ indicates is available.

Let us now consider the case where the Q-design is of the $L_1 \times L_2 \times \cdots \times L_F$ type. Denote by $q_f(\pi_j)$ the vector

$$(q_{f1}(\pi_j), q_{f2}(\pi_j) ,\ldots, q_{fL_f}(\pi_j)) ,$$

$q_{f\ell}(\pi_j)$ being the marginal Q-score of the j-th subject for the f-th factor at the ℓ-th level. If we adopt the restriction that holds the scaling function to be skew symmetric about $(I+1)/2$ (the palindrome axiom), then, for each f,

$$\sum_{\ell=1}^{L_f} q_{f\ell}(\pi_j) = 0 \ . \qquad \text{II.B.1(2)}$$

This redundancy requires that $D = \Sigma_f\ (L_f-1)$, if Q is to be of full rank D. This is not convenient, however, so we shall use the redundant item design matrix Q with $D = \Sigma_f L_f$, but of rank $\Sigma_f(L_f-1)$.

We generalize the conditional likelihood of the 2^F balanced factorial design by considering all permutations of the elements in $q_f(\pi_j)$. Thus,

$$e^{\mathcal{L}_{BC}(\beta)} = \prod_{j=1}^{J} \left\{ \frac{\exp\{q(\pi_j)\beta w_j'\}}{\prod_{f=1}^{F} \Sigma_{P_f} \exp\{q_f(\pi_j)P_f\beta_f w_j'\}} \right\} \qquad \text{II.B.1(3)}$$

where $\beta = (\beta_1,\ldots,\beta_f,\ldots,\beta_F)$ and $q(\pi_j) = (q_1(\pi_j)\ ,\ldots,\ q_f(\pi_j),\ \ldots,\ q_F(\pi_j))$. Σ_{P_f} denotes summation over all permutation matrices of order L_f. Because of the restriction II.B.1(2), the constraints

$$\sum_{\ell=1}^{L_f} \beta_{f\ell} = 0 \ ,$$

where $\beta_{f\ell}$ are the columns of β_f, are necessary for β to be identified. The utility of II.B.1(3) is that the denominator is computable.

As for the 2^F design, a geometric interpretation of the orbit of conditioning is possible here. For example, when $F = 2$, $L_1 = 2$ and $L_2 = 3$, the vertices that compose the orbit of the

conditioning correspond to the corners of a right six-sided prism whose two hexagonal faces are generated by the permutations of the three levels within the second factor, each of the two faces corresponding to a level of the first factor. As for the 2^F design, the vertices

$$\{q(\pi_j)P : P \text{ a permutation of levels within factors}\}$$

compose an orbit whose centroid is the origin. Also, if the elements $q_f(\pi_j)$ are not identical for $f = 1,\ldots,F$, these vertices form a basis which spans the whole design space.

The information that is suppressed by conditioning as above deserves discussion. Key is the observation that the design space is spanned only when the elements within each $q_f(\pi_j)$ are not identical. Suppose that for all subjects that the elements $q_{f_0}(\pi_j)$ are identical (zero). While one might wish the corresponding matrix β_{f_0} to be estimated by a null array, in fact β_{f_0} is not identifiable in $e^{\mathcal{L}_{BC}(\beta)}$. If, now, only the j_0-th subject has non-identical $q_{f_0}(\pi_j)$, one would anticipate that the preponderance of null $q_{f_0}(\pi_j)$'s would push any reasonable estimate of β_{f_0} toward the null array. In fact, in minimizing $\mathcal{L}_{BC}(\beta)$ with respect to β_{f_0}, only subject j_0's scores would contribute; adding additional subjects with $q_{f_0}(\pi_j) = 0$ would not dilute the effect that subject j_0's scores would have on the estimate of β_{f_0}, while deleting subject j_0 would make β_{f_0} unidentifiable. For this reason, one might characterize the information upon which $e^{\mathcal{L}_{BC}(\beta)}$ is conditional as the "magnitudes of the effects," that

the information that remains is in the "directions of the effects." Of course, this description lacks precision; nevertheless, the manner in which a subject possessing especially large Q-scores on a particular factor can sway the conditional maximum likelihood estimates provides one motivation for considering other kinds of conditional likelihoods.

II.B.2. The unbalanced Q-set.

The conditional likelihood in II.B.1(2) can be unsatisfactory for two reasons. First, the likelihood $e^{\mathcal{L}_{BC}(\beta)}$ is only possible in principle when the Q-set is structured as a completely balanced, cross-factorial design. Not only are certain Q-sets not completely of this type, lacking balance for instance, but often even for such carefully designed Q-sets one desires to append to the item design matrix certain nuisance covariates, e.g. measures of the items' social desirability, concept complexity, vocabulary level, etc. (Sundland [1962]) or to consider interactions among the main factors. Thus, a conditional likelihood applicable to an arbitrary item design matrix is desirable. Second, as noted in the previous section, for some data sets with certain Q-matrices, $e^{\mathcal{L}_{BC}(\beta)}$ can behave unsatisfactorially, by over emphasizing the contributions of certain subjects. In these same instances, numerical problems in the maximization process may result.

In this section we shall develop a conditional likelihood that makes use of a representation that is dual to that of a permutation.

<u>Definition</u>. Let σ, a <u>ranking</u>, represent the ordering of a Q-deck such that

$$\sigma(i) = \text{rank of the i-th item in the Q-sort.}$$

Note that for a given Q-sort represented by σ and π, $\sigma(\pi(k)) = k$ and $\pi(\sigma(i)) = i$. So σ and π are inverses of one another. We shall denote the range of σ by $N(I)$, and refer to it as the *rankings space*.

Definition. A *shuffle*, $w{:}N(I) \rightarrow N(I)$, is a one-to-one onto map on the ranking space $N(I)$ that can be represented by a function $\omega\colon \{1,2,\ldots,I\} \rightarrow \{1,2,\ldots,I\}$ that is one-to-one and onto such that

$$w(\sigma)(i) = \omega(\sigma(i)) \ .$$

We refer to ω as the *shuffler*.

A shuffle operates to rearrange the ordering of a Q-set in a particular, systematic way. One can visualize a shuffle as an automatic card shuffling machine. Its argument is the input Q-sort; its result is a reshuffled deck. More importantly, a shuffle ignores the indices $\{i\}$ of the cards and operates only on their ranks.

Definition. The *composition* ($\circ$) of two shuffles, w_1 and w_2, with corresponding shufflers ω_1 and ω_2, is such that

$$w_1 \circ w_2(\sigma)(i) = \omega_1(\omega_2(\sigma(i))) = w_1(w_2(\sigma))(i) \ .$$

By this definition, the sorted deck $w_1 \circ w_2(\sigma)$ results from shuffling σ by w_2 and then shuffling that result by w_1.

Note that each shuffler ω induces a semi-group on $N(I)$ whose corresponding shuffle is the generator. This brings us to our next construct.

__Definition__. Let Ω be a set of shufflers whose corresponding set of shuffles is closed under composition. A __random shuffler__ is the probability space whose sample space is Ω that assigns equal probability to each element of Ω.

Intuitively, Ω could be seen as an advancement in card shuffling technology over the simple shufflers, $\{\omega\}$. Such a machine has at its disposal several shufflers, and for any particular task it chooses one of these shufflers at random. More importantly, if one recycles the output through the random shuffling mechanism once again, the probability distribution induced on $N(I)$ remains unchanged.

This set of definitions is now used to simplify the full likelihood. We propose to condition on a random shuffler. By this scheme, any $p_j(\pi_j)$ is then of the form

$$p_j(\pi_j) = \frac{\exp\{[\Sigma_{k=1}^{I} S(k)Q_{\pi_j(k)}]\beta\, w_j'\}}{\Sigma_{\omega\,\varepsilon\,\Omega} \exp\{[\Sigma_{k'=1}^{I} S(\omega(k'))\, Q_{\pi_j(k')}]\beta\, w_j'\}} \qquad \text{II.B.2(1)}$$

The terms $\{\Sigma_{k'=1}^{I} S(\omega(k))Q_{\pi_j(k)} : \omega\,\varepsilon\,\Omega\}$ are then the orbits with respect to which the conditioning is made. To emphasize the central role Ω plays in the creation of these orbits, and to reinforce the sobriety of our endeavor, Ω, the sample space of the random shuffler shall be termed the __orbit generator__ in all of the following.

We now turn our attention to specific properties desired of Ω.

__Definition__. An orbit generator Ω is unbiased with respect to S if

$$\frac{1}{\#\Omega} \sum_{\omega\in\Omega} S(\omega(r)) = 0 \quad \text{for} \quad r = 1,2,\ldots,I\ ,$$

where $\# \Omega$ denotes the number of elements of Ω. Thus, Ω is unbiased if the centroid of the scores $S(\omega(r))$ is the same as the mean of the scaling function $S(\cdot)$.

<u>Definition</u>. An orbit generator Ω is said to contain its reversals if

$$\text{whenever } \omega \in \Omega \text{ then } \bar{\omega} \in \Omega ,$$

where $\bar{\omega}(r) = \omega(I-r+1)$. Clearly any orbit that contains its reversals is unbiased with respect to skew-symmetric scaling functions.

The smallest unbiased orbit generator for skew-symmetric $S(\cdot)$ is $\Omega_0 = \{e,\bar{e}\}$, where e is the identity shuffler and $\bar{e}$ is its reverse ($e(r) = r$, $\bar{e}(r) = I-r+1$). A natural criticism of Ω_0 is that it is too sparse: Consider the vectors S_ω, such that $S_\omega(r) = S(\omega(r))$. Then of the $(I-1)$-dimensional space in which resides the $\{S_\omega\text{: all } \omega\}$, Ω_0 spans only a one-dimensional subspace. With this objection in mind, the following classes of generators are proposed.

For notational convenience, rather than labeling the ranks by the numbers $1,\ldots,I$, we label them by the numbers $0,1,\ldots,I-1$. Also we denote a shuffler explicitly by an I-tuple. Thus, $\omega = (\omega(0),\omega(1),\ldots,\omega(I))$, and $\omega(r)$ connotes the rank to which the shuffler ω moves the item that bore rank r.

Consider the following table:

	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7
2	0	2	4	6	0	2	4	6
3	0	3	6	1	4	7	2	5
4	0	4	0	4	0	4	0	4
5	0	5	2	7	4	1	6	3
6	0	6	4	2	0	6	4	2
7	0	7	6	5	4	3	2	1

This table is the multiplication table modulo 8. Note also that rows 1, 3, 5, and 7 are permutations (or shufflers), while rows 0, 2, 4, and 6 repeat digits and so are not shufflers. (Not coincidentally, 0, 2, 4, and 6 have factors common to 8.) For each shuffler from this multiplication, that is, for each of

0	1	2	3	4	5	6	7
0	3	6	1	4	7	2	5
0	5	2	7	4	1	6	3
0	7	6	5	4	3	2	1

consider its cyclics. For 0 1 2 3 4 5 6 7 they are

0	1	2	3	4	5	6	7
7	0	1	2	3	4	5	6
6	7	0	1	2	3	4	5
5	6	7	0	1	2	3	4
4	5	6	7	0	1	2	3
3	4	5	6	7	0	1	2
2	3	4	5	6	7	0	1
1	2	3	4	5	6	7	0

Notationally, the cyclics of a shuffler ω are the compositions ω, $c\omega$,

$c c\omega = c^2\omega, \ldots, c^{I-1}\omega$ where $c = (1\ 2\ 3\ \cdots\ I\text{-}1\ 0)$. Intuitively, the shuffler c takes the item at the top of the deck and places it on the bottom.

The shuffles that are the rows embedded in the multiplication table modulo I are, in general,

$$\omega_p = (0,\ 1\cdot p,\ 2\cdot p\ , \ldots,\ (I-1)\cdot p)$$

where p is relatively prime with respect to I and where the operation "$\cdot$" is multiplication modulo I. Define Ω_p by

$$\Omega_p = \{\omega\colon \omega = c^k\omega_p,\ k=0,1,\ldots,I-1\}\ ;$$

thus Ω_p is the set of all cyclics of ω_p. Finally, define

$$\Omega' = U'_p\ \Omega_p$$

where U'_p denotes union over all indices p relatively prime to I.

<u>Proposition</u>. Ω' is closed with respect to the composition operation. Also, $\Sigma' = \{S_\omega\colon \omega\epsilon\Omega'\}$ spans the (I-1)-dimensional linear subspace that holds the full lattice $\{S_\omega\colon$ all $\omega\}$.

Geometrically, Ω' being closed with respect to composition translates to the elements of Σ' being vertices of some regular polyhedron in the (I-1)-dimensional space of all rankings. That the elements of Σ' span this space implies that this polyhedron is "solid" in the (I-1)-dimensonal space. With these two properties,

Σ' intuitively satisfies the requirement that the chosen subset of $\{S_\omega\}$ be evenly scattered over the surface of the (I-1)-sphere. We shall call Ω' the <u>cyclic prime orbit generator</u>, and call the likelihood Ω' induces in II.B.2(1) the cyclic prime likelihood, denoted $_e\mathcal{L}_{cp}(\beta)$.

The above proposition has the following generalization. Let Π be a subset of the relative primes of I such that if p_1 and p_2 are elements of Π, so also is p_1p_2 (multiplication modulo I). Unity is also an element of Π. Define ω_p as before by

$$\omega_p = (0, 1p, 2p, \ldots, (I-1)p) \quad \text{for all} \quad p \in \Pi ,$$

and define Ω_{pq}, where q is any number which shares a factor of I (q may equal unity), by

$$\Omega_{pq} = \{\omega: \ \omega = c^{kq}\omega_p, \ k = 0,1,2,\ldots,I-1\} .$$

Then the following generalization is true.

<u>Proposition</u>. $\Omega(\Pi, q) = \cup_{p\in\Pi}\,\Omega_{pq}$ is closed with respect to the composition operation.

This proposition allows us to consider orbits smaller than those induced by Ω'.

One should note that in simplifying the likelihood to this conditional form, an element of arbitrariness has been introduced that was not present previously. Whereas the full likelihood was invariant to the manner in which ties were broken, as was the balanced, cross-factorial likelihood, those likelihoods based upon

orbit generators are typically not so invariant. Unfortunately, to enforce this invariance on this class of likelihoods would add considerable computational effort, running counter to the primary motivation for considering this class of estimates.

III. Asymptotic properties of likelihood analyses in conditional problems.

III.A. Maximum likelihood estimation and maximum conditional likelihood estimation.

All results of consistency, uniqueness, and asymptotic normality follow from specialization of results due Andersen (1970), and are particularly easy because $p_j(\cdot)$ is an exponential family parametrized by β.

III.A.1. Consistency and uniqueness.

In chapter II, focus was exclusively upon various likelihood functions. This focus may seem curious to some, for the likelihood functions do not furnish us directly with either estimates of β or with inferential procedures. We now close this gap. Estimates of β are obtained by using that value of β that maximizes whichever likelihood is convenient. The estimation equations are, for the full likelihood,

$$\nabla_\beta \mathcal{L}(\beta) = 0 = \sum_{j=1}^{J} \{q(\pi_j)w_j' - \mathcal{E}_\pi\{q(\pi)w_j'|\beta\}\}$$

where $\mathcal{E}_\pi\{q(\pi)w_j'|\beta\} = \Sigma_\pi\, p_j(\pi)\, q(\pi)w_j'$, the expectation relative to the distribution induced by $p_j(\pi)$.

The various conditional likelihoods have estimation equations of a similar form. For example, the cyclic prime likelihood has the equations

$$\nabla_\beta \mathcal{L}_{cp}(\beta) = 0 = \sum_{j=1}^{J} \{q(\pi_j)w_j' - \mathcal{E}_{\pi,cp}\{q(\pi)w_j'|\beta\}\}$$

where $\mathcal{E}_{\pi,cp}\{\cdot\}$ denotes the conditional expectation given the cyclic prime orbit.

The second derivatives are of the form

$$\nabla_\beta \mathcal{L}(\beta)\nabla_\beta' = -\sum_{j=1}^{J} \mathrm{Cov}_{\mathcal{L}} \{q(\pi)w_j', \; w_j q(\pi)'\}$$

from which we may infer that the maximum is achieved and, if well-identified, unique, due to the positive definiteness of $-\nabla_\beta \mathcal{L}(\beta)\nabla_\beta'$. (Also, if $-\nabla_\beta \mathcal{L}(\beta)\nabla_\beta'$ is only positive <u>semi</u>-definite, then β is not fully identified; constraining β in the proper way will identify it, whence $\mathcal{L}(\beta)$ has a unique constrained maximum.)

For reasons related directly to the uniqueness of the solution to the maximum (conditional) likelihood equations, it follows that as J becomes large, the solution, $\hat{\beta}$, converges to the true value β.

III.A.2. *Asymptotic normality.*

Let us denote the Fisher information matrix with respect to the likelihood $\mathcal{L}(\cdot)$ by $I_{\mathcal{L}}(\beta)$ and define it by

$$J I_{\mathcal{L}}^{-1}(\beta) = -\nabla_\beta \mathcal{L}(\beta)\nabla_\beta' = \sum_{j=1}^{J} \mathrm{Cov}_{\mathcal{L}}\{q(\pi)w_j', w_j q(\pi)'\}$$

where $\mathcal{L}$ is whichever (conditional) likelihood is being employed and where $\mathrm{Cov}_{\mathcal{L}}\{\cdot,\cdot\}$ is the covariance with respect to $\mathcal{L}$.

Then from Andersen (1970), it follows that $J^{1/2}(\hat{\beta}-\beta)$ is distributed $\mathcal{N}(0, I_{\mathcal{L}}(\beta))$ as J become large; simple inferences may be made on this basis.

III.B. Generalized likelihood ratio hypothesis testing.

While the property of $\hat{\beta}$ being asymptotically normal may be exploited directly to form tests of hypotheses, the generalized likelihood ratio statistic (glr) is usually more convenient. This is because the glr is available as a direct consequence of the maximizing algorithm. The general form of this statistic is as follows.

Consider the hypotheses

$$H_0: \beta \varepsilon \mathcal{H}_0 \quad \text{vs} \quad H_1: \beta \varepsilon \mathcal{H}_1 .$$

The conditions which we adopt are that (a) $\mathcal{H}_0$ is a subset of $\mathcal{H}_1$ and (b) $\mathcal{H}_0$ contains no subset that is an open set in $\mathcal{H}_1$. Under these conditions, and for our model the statistic

$$\text{glr}(01) = \frac{\max\limits_{\beta\varepsilon \mathcal{H}_0} \exp\{\mathcal{L}(\beta)\}}{\max\limits_{\beta\varepsilon \mathcal{H}_1} \exp\{\mathcal{L}(\beta)\}}$$

has nice properties. In particular, the asymptotic distribution may be derived by using the fact referenced in the previous section that $\hat{\beta}$ is normally distributed. The asymptotic distribution of glr(01) is such that -2 log glr(01) is approximately χ^2 with D_1-D_0 degrees of freedom, where D_i is the dimensionality of hypothesis $\mathcal{H}_i$. We shall use the above result for the solution of the hypothesis testing problems posed in the next chapter.

IV. Problems of applied interest.

This chapter considers two kinds of applied problems that naturally arise in the context of the structured Q-sort. The first of these, the testing of nested hypotheses to assess the significance of sets of β-coefficients, has a particularly simple form. The simplicity of this theory results not from any special properties of the Q-sort model but from the well-known theory of (conditional) generalized likelihood-ratio tests. An example is considered in detail to establish the appropriateness of this theory.

In the latter part of this chapter, a class of hypotheses is described that falls outside the natural domain of the generalized likelihood ratio tests. Precisely because these hypotheses are central to Stephenson's structured Q-sort methodology, special attention is required to develop an appropriate test.

IV.A. Testing nested hypotheses.

Before proceeding, the following change in notation is convenient. In chapter III, $\mathcal{H}_\ell$, $\ell = 0,1,\ldots,$ denoted subspaces of the parameter space to which belonged β. We replace this convention with another. Hereafter, let $\{\mathcal{H}_\ell\}$ denote sets of index pairs. Thus, a hypothesis H_ℓ can be of the form

$$H_\ell :\ \beta_{dk} = 0 \quad \text{for all} \quad dk \in \mathcal{H}_\ell \ .$$

Although these classes of hypotheses are less general than those of chapter III, they are sufficient for most practical problems.

The general problem upon which we focus is the test of the hypothesis

$$H_0: \beta_{dk} = 0 \quad \text{for} \quad dk \in \mathcal{H}_0$$

versus

$$H_1: \beta_{dk} = 0 \quad \text{for} \quad dk \in \mathcal{H}_1 ,$$

where the key regularity condition is that $\mathcal{H}_1$ is a proper subset of $\mathcal{H}_0$. If we let $g = \# \mathcal{H}_0 - \# \mathcal{H}_1$, this condition is sufficient to show that $-2 \log \lambda_{01}$ is asymptotically distributed as a chi-square with g degrees of freedom, where λ_{01} is defined by

$$\frac{\sup\{\exp\{\mathcal{L}_{cp}(\beta)\}: \beta_{dk} = 0,\ dk \in \mathcal{H}_0\}}{\sup\{\exp\{\mathcal{L}_{cp}(\beta)\}: \beta_{dk} = 0,\ dk \in \mathcal{H}_1\}} .$$

One example for which such a hypothesis might be formulated is the following:

Very often in the development of a Q-set, the matrix Q is chosen and fixed; only then are the Q-items formulated. The most notable instances of this procedure are the balanced, cross-factorial designs Stephenson built into his Q-sets. A primary criticism of this procedure, articulated by Sundland (1962), is that no guarantee can be made that the items are actually sorted in response to those properties that led to their choice in the first place. In particular, the rater may be reacting primarily to the social desirability of the items, or their conceptual complexity, or an interaction between such features. Because of these concerns, it

may be desirable to augment the item design matrix with additional columns that represent such "nuisance" covariates. Once incorporated, the coefficients of these covariates can then be evaluated to determine if they are significantly associated with the propensity of any item to being ranked highly. The form of the null hypothesis is

$$H_0\colon\ \beta_{dk} = 0, \quad d = D_1,\ldots,D_2 \quad \text{and all} \quad k$$

where the rows $D_1,\ldots,D_2$ would represent such "nuisance" covariates; the natural alterantive hypothesis is one where $\mathcal{H}_1$ is an empty index set.

The above null hypothesis may lack power by being too global. While the raters may be responding to an unintended concomitant feature of the items, they may be less likely, if well-trained, to confound the error by reacting to this feature in different degrees with different subjects. Motivated by this consideration, a hypothesis intermediate between H_0 and the general alternative of the form

$$H'\colon \beta_{dk} = 0, \quad d = D_1,\ldots,D_2 \quad \text{and} \quad k = 1,$$

(where $w_{j1} = 1$, the constant part of the predictor space), could represent an <u>a priori</u> direction that would successfully concentrate the power of test of H_0. In this scheme, the sequence of hypothesis tests H_0 vs. H' and then H' vs. H_1 provides a stepwise procedure with the potential of greater acuity than that which would result from simply testing H_0 vs. H_1 globally. (An incidental but convenient property of this form of the stepwise procedure is that the two tests of which it is composed are independent.)

IV.B. Testing validity by using a design that stratifies subjects.

In the previous section, the standard theory of generalized likelihood ratio testing was sketched. Recall, however, that the example included consisted of the consideration of a relatively peripheral issue — the significance of nuisance covariates. The choice of this example is not accidental. The central hypotheses of Q-studies are usually not of the general form of the previous section; rather the sign of the coefficients is usually specified in the alternative. This is because the samples of structured Q-sort studies are often configured by deliberately choosing certain kinds of subjects. Stephenson proposed selecting individuals with characteristics that could be theoretically predicted. If their Q-sorts did not correspond well with the predictions of the theory, the theory was considered invalidated.

As an example of this, consider a Q-set representing the typology of Spranger (1928) that partitioned persons into the types: religious political, theoretical, economic, aesthetic, and social. The idea is then to test this Q-set upon clerics, whose value systems one would expect to be religious, politicians, whose value systems one would expect to be political, academics, bankers, artists and bartenders, each subject to a natural a priori classification. If the Q-sort shows clerics not being particularly religious, politicians not particularly political, and so on, then the most natural conclusion is that the Q-set, the instrument, is no good, and that by inference, the theory the Q-set represents is invalid.

Formally, this problem may be presented by hypotheses of the form

$$H_0;\ \beta_{dk} > 0,\ dk \in \mathcal{H} \text{ vs. } H_1:\ \beta_{dk} \text{ arbitrary, } dk \in \mathcal{H}\ ,$$

where $\mathcal{H}$ is a designated index set. The generalized likelihood ratio test is not appropriate as it stands because the dimensionality of the null and alternative hypotheses is the same.

The following modification makes generalized likelihood ratio hypothesis testing feasible: Under both H_0' and H_1' restrict $\beta_{dk} = \theta$, $dk \in \mathcal{H}$, and θ unknown. Then with

$$H_0'\colon \beta_{dk} = \theta,\ \ dk \in \mathcal{H},\ \theta \geq 0 \ \text{ vs } \ H_1'\colon \beta_{dk} = \theta,\ dk \in \mathcal{H},\ \ \theta < 0\ ,$$

we obtain a structure where an assessment of this one-sided hypothesis is easily made.

The reparametrizing of H_0' and H_1' says that the degree to which a cleric is religious is the same as the degree to which a politician is political and the degree to which a banker is economic. This seems not unreasonable. One may wish to test this hypothesis, however, and the following test is independent of the latter. Let H_0'' and H_1'' be

$$H_0''\colon \beta_{dk} = \theta,\ dk \in \mathcal{H},\ \ -\infty < \theta < \infty \ \text{ vs. } \ H_1''\colon \ \beta_{dk} \text{ arbitrary.}$$

In practice one might wish to test H_0'' and, if accepted, test H_0'. Rejection of either case can be construed as evidence against validity. The interpretation of the rejection differs between the two cases,

however. H_0'' postulates the lack of interactions that might otherwise confound the test of the main effect; its rejection would imply the presence of such interactions. H_0' postulates the direction of the main effect; its rejection invalidates the theory that was the basis of the construction of the Q-set.

V. Analysis of the unstructured Q-sort.

In chapters II, III, and IV, the item design matrix Q and the subject design matrix W were always assumed known. This was the case of the structured Q-sort. The Q-set had an a priori structure, as did the subjects; the problem was the manner in which these two structures related. The estimate of the matrix β represented a solution to this problem. In this chapter at least one of the design matrices, Q or W, is unknown and some reasonable estimates of them is desired. The form of β, on the other hand, is no longer of interest; indeed, because it is unidentifiable, the issue of its form is moot.

V.A. The statement of the problem and its algorithm.

Chapter II reparametrized $\rho(i,j)$ to be of the form

$$\rho(i,j) = \exp\{Q_i \beta w_j'\}, \quad i = 1,\ldots,I;\ j = 1,\ldots,J$$

where Q_i and w_j were known row vectors and β unknown. In the present section both Q_i and w_j are unknown. β, no longer identifiable, is suppressed; as a result $\rho(i,j)$ can be taken to be of the form

$$\rho(i,j) = \exp\{\sum_{d=1}^{D} q_{id}\, w_{jd}\} = \exp\{Q_i w_j'\} \qquad \text{V.A. (1)}$$

where q_{id} and w_{jd} are unknown scalars. This formulation is very similar to that of the factor analysis problem: the $\{q_{id}\}$ are the items' factor loadings and the $\{w_{jd}\}$ are the subjects' factor scores. The absence of the matrix β can be attributed, by this analogy, to the indeterminate linear transformation that may be shunted between the factor loadings and the factor scores.

The $\{\rho(i,j)\}$ in V.A.(1) are not fully identified. If we define

$$\rho(i,j) = \exp\{Q_i \Lambda w_j'\}$$

with the restrictions that

$$\begin{aligned} &\Sigma_i Q_{id} = 0, \quad \text{for all} \quad d = 1,\dots,D \\ &\frac{1}{I} Q'Q = I_D \quad \text{and} \quad \frac{1}{J} W'W = I_K \end{aligned} \qquad \text{V.A.(2)}$$

and $\Lambda = \mathrm{diag}(\lambda_1,\dots,\lambda_D)$, and require $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_D$ then the parameters Q_i, Λ, and w_j are essentially identified. (The remaining ambiguity takes place only when some of the λ_d's are equal.)

The algorithm seeks to maximize the objective function

$$\mathcal{L}_{cp}(Q \Lambda W), \text{ subject to } \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_D \text{ and V.A.(2)} \qquad \text{V.A.(3)}$$

where $\mathcal{L}_{cp}(Q \Lambda W)$ is such that

$$e^{\mathcal{L}_{cp}(Q \wedge W)} = \prod_{j=1}^{J} \left\{ \frac{\exp\{[\sum_k S(k) Q_{\pi_j(k)}] \wedge w_j'\}}{\sum_{\omega \in \Omega'} \exp\{[\sum_k S(\omega(k)) Q_{\pi_j(k)}] \wedge w_j'\}} \right\}$$

The maximization of $\mathcal{L}_{cp}$ over such a high dimensional space (the dimension is $D(I+J-1)$ is impractical for a moderate number of subjects. When $I = 100$, a typical number, Ω' has four thousand orbits. For example, if $J = 100$, in order to evaluate $\mathcal{L}_{cp}$ even once, four hundred thousand orbits would need be evaluated. This is not generally practical.

One natural modification is to consider the smaller orbits $\Omega(\Pi,q)$ that were described in II.B.2. If it is desired that $\Omega(\Pi,q)$ span (I-1)-dimensional space of rankings, then $\#\Omega(\Pi,q)$ needs to have at least I elements. It may be desired that $\Omega(\Pi,q)$ contain its own reversals; then $\Omega(\Pi,q)$ must contain at least $2I$ elements.

V.B. Variations to the analysis of the unstructured Q-sort.

Two variations of the above analysis may be posed. (1) Rather than having both design matrices unknown, only one design matrix may be unknown; the other is specified. (2) Having estimated one or both of the matrices Q and W, rotations to simple structure are often desirable in order to ease the interpretation of the factor structure.

V.B.1. The factor analysis given one specified design matrix.

Because the problem of estimating the subject design matrix when the item design matrix is known is closely parallel to that of estimating the item design matrix when the subject design matrix is specified, only the latter will be discussed.

Since $\frac{1}{J} W'W$ need not be the identity matrix consider the transformation T so that $\frac{1}{J}(TW)'(TW) = I$. Let $X = TW$.

The objective function is then

$$\mathcal{L}(Q\ \Lambda\ X') \text{ subject to } \frac{1}{I} Q'Q = I;\ \sum_i Q_{id} = 0, \text{ for all } d; \quad \text{and} \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_D \geq 0 \tag{V.B.1(1)}$$

and we maximize V.B.1(1) with respect to Q and Λ and use the maximizing values as the magnitudes and directions of the factor structure. $\beta = \Lambda T'^{-1}$ may be interpreted as the parameter matrix relating the factor loadings Q to the "factor scores" W.

V.B.2. Rotations to simple structure.

The restrictions V.A.(2) are only necessary for technical reasons — to ensure that Q, Λ, and W are identifiable. These estimates need not be easy to interpret, however; one may wish to exploit their lack of definition by selecting orthogonal rotations of Q and W to make their structure more comprehensible. In the mid fifties several criteria for simple structure that furnish precise algorithms were proposed (Ferguson [1954], Carroll [1953], Neuhaus and Wrigley [1954], Saunders [1953], and Kaiser [1956]).

Each of these rotations to simple structure operates on the factor loadings. As a result, the factor loading matrix Q (and its dual, the factor score matrix W) is in the correct form to be rotated by VARIMAX, QUARTIMAX, or whatever; that the factor loading matrix is the result of maximizing a cyclic prime likelihood is not relevant.

However, the matrix Λ needs to be transformed if either Q or W are rotated. If $Q(R) = QR$ is the result of the rotation R, and if $W(S) = WS$ is the result of the rotation S, then Λ needs to be replaced by $\beta = R' \Lambda S$; thus

$$\log \rho(i,j) = (Q \Lambda W')_{ij} = Q(R)(R' \Lambda S) W(S)'_{ij} .$$

Therefore, the simplicity of any rotations of Q or W needs to be weighed against the complexity such transformations may induce on the matrix β. Incorporating such an index of the simplicity of β into the algorithm optimizating the simplicity of Q and W seems appropriate.

VI. An example.

The primary consideration in developing a class of likelihood models conditional on random shufflers is computational feasibility. Even so, maximizing conditional likelihood functions remains CPU intensive; just how intensive is best illustrated by a practical example.

VI.A. Description of origin of the data.

The data was provided by Phyllis Sherlock, PhD; the reader is referred to Sherlock (1980) for the substantive details of the origin of the data. For the present purpose of providing a practical example of the analysis, the following summary of Sherlock's design of the Q-set is provided.

(1) The theoretical background for the structure of the Q-set is the typology of female psychologies of Toni Wolff, who developed it in the framework of the analytical psychology of Carl Jung. Four types of psychologies are postulated: The Mother, the Amazon, the Hetaira, and the Medium. These four types are arranged as two bipolar pairs - the "Mother-Hetaira" and the "Amazon-Medium"; these two bipolarities compose a coordinate system of two orthogonal axes. See Figure 1.

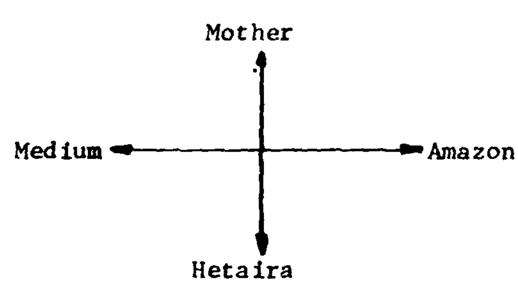

Figure 1.

The point at which these axes intersect is the origin in this coordinate space, the bipolarities are the coordinate directions. In this theoretical model, the psychology of any particular female corresponds to a point in this two-dimensional space.

(2) Based on Wolff's descriptions of each of these four typologies, Sherlock developed a set of Q-items, each item consisting of simple adjectives or short phrases. For each typology, Sherlock associated twenty-five Q-items; the Q-set was composed of these four groups of twenty-five items - one hundred items in all.

The item design matrix, Q, was generated in the following way:

(3) Sherlock had four experts in Wolff's typology each sort the Q-set four times, one sort for each of the female types. The experts were told to sort for the <u>ideal</u> Mother, the <u>ideal</u> Amazon, and so on.

(4) Based on this expert data, the item design matrix was constructed as follows:

(a) Within each expert, the four scores of each item were centered to have mean zero across the four conditions (the conditions of sorting for the ideal Mother, the ideal Amazon,...).

(b) For each condition, the scores were averaged across raters, giving a design matrix of rank four.

(c) Each column of this matrix was then "centered", that is, made orthogonal to the 1 x 100 vector, each of whose elements is unity. The resulting design matrix had four columns but had rank three.

(d) This design matrix was in turn transformed to reflect the theoretical coordinate structure. By subtracting the design column corresponding to the Hetaira type from the column corresponding to the Mother type, a design column representing the ordinate (Mother-Hetaira) was produced. By subtracting the column corresponding to the Medium type from that corresponding to the Amazon type, a design column representing the abscissa was formed. (The design could have been saturated by including the column consisting of Mother + Hetaira - Amazon - Medium, but for simplicty this was not done.)

The subject design matrix, W, was chosen as follows:

(5) A constant covariate was included to reflect the overall propensity for the subjects sampled to by any particular type.

(6) Two scores from the Meyers-Briggs inventory were included to reflect some of Sherlock's hypotheses. The Meyers-Briggs is a paper and pencil type questionnaire designed to measure four traits central to Jung's personality theory. The traits are: extraversion-introversion (E-I), thinking-feeling (T-F), sensation-intuition (S-N), and judging-perceiving (J-P). The scores E-I and J-P, with the intercept, made up the subject design matrix.

(7) Sherlock collected Q-sorts from 80 individuals. Of these, three were rejected from the computer runs because of coding aberrations. Therefore, the analysis below is based on 77 Q-sorts.

VI.B. Description of the technical characteristics of the algorithm.

The analysis presented below was implemented on the IBM 3033 computer located at Stanford's Center for Information Technology. The source code was written in FORTRAN and compiled at level G. Several IMSL routines were employed to perform some of the standard transformations of matrices required. The algorithm is a "protected" Newton-Raphson iterative scheme, modified to ensure that each refinement brings an increase in the likelihood.

For the data set described in section VI.A, each iteration, consisting of an evaluation of its likelihood, its gradient, and its matrix of partials takes approximately 0.20 CPU minutes. Eleven iterations were required to locate the solution reported below. The default allocation of 256 Kilobytes of core memory was adequate.

The orbit generator employed consisted of the 200 shufflers whose three generators are the following:

$$\omega_{13} = (0,\ 13,\ 26, \ldots,\ 87)\ ,$$

that is, $\omega_{13}(r) = 13 \cdot r \pmod{100}$,

$$\bar{e} = (99,\ 98,\ 97, \ldots, 1,\ 0)\ ,$$

and

$$c^{20} = (20,\ 21,\ 22, \ldots, 99,\ 0,\ 1,\ 2, \ldots, 19)\ .$$

The group generated by ω_{13} has twenty elements, that generated by $\bar{e}$ two elements, and that generated by c^{20} five, for a total of $20 \times 2 \times 5 = 200$ elements.

VI.C. Description of the results.

In maximizing the likelihood, its logarithm went from -407.97 at $\beta = 0$ to -407.56 at its maximum, a change corresponding to a chi-square with six degrees of freedom of 0.82. From this one may conclude that the model fitted did not significantly differ from the $\beta = 0$ model. One should add, however, that Sherlock had no strong hypotheses about the relation of Wolff's typology to either of the two scores from the Meyer-Briggs; the lack of any significant effects has, therefore, no particular impact on the validity of Wolff's typology.

The format of the answers that this likelihood model produces should be of interest to any who seek to build predictive models of structured Q-sort data. Essentially this format has three features:

(1) Coefficients are fit in a manner that allows them to be interpreted as regression coefficients; they may be standardized. See Table 1.

(2) The coefficients allow a pair of dual visualizations: The underlying dimensions of the item design space may be represented as coordinates in the subject design space. Similarly the underlying dimensions of the subject design space may be represented as coordinates in the item design space. See Figures 2 and 3. This is the duality described in section II.A.1.

(3) The variances and correlations of the coefficients are obtained from the Fisher information matrix. See Table 2.

Table 1. Coefficients of the Sherlock Data

	Intercept	Extraversion-Introversion	Judging-Perceiving
Mother-Hetaira	$-2.20 \times 10^{-4}(-3 \times 10^{-4})$	$-2.09 \times 10^{-3}(0.012)$	$-7.51 \times 10^{-3}(-.043)$
Amazon-Medium	$2.25 \times 10^{-4}(3 \times 10^{-4})$	$1.97 \times 10^{-3}(0.011)$	$7.71 \times 10^{-3}(0.044)$

coefficient (standardized coefficient)

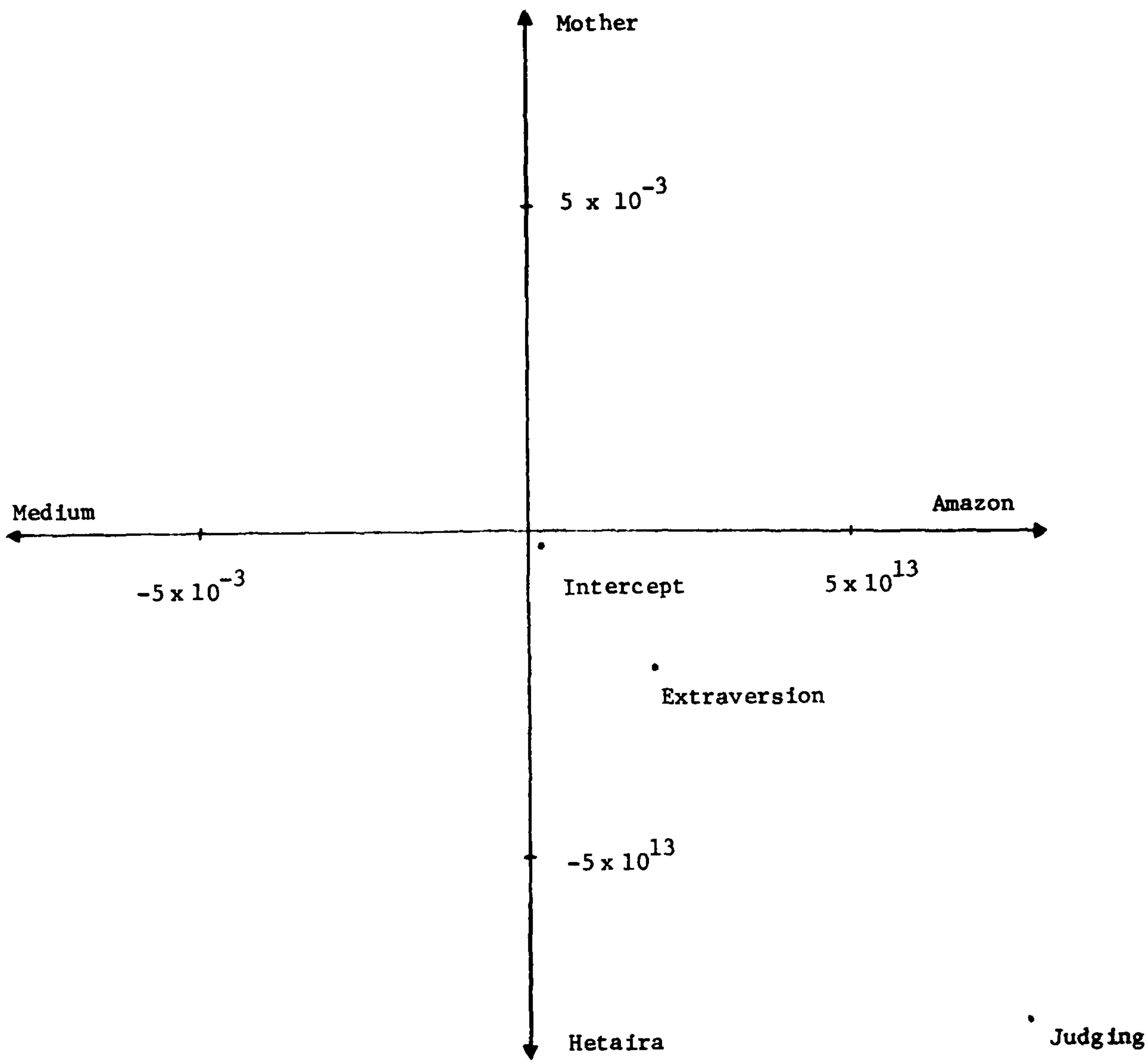

Figure 2. Location of the covariates of the subject design matrix in the item design space.

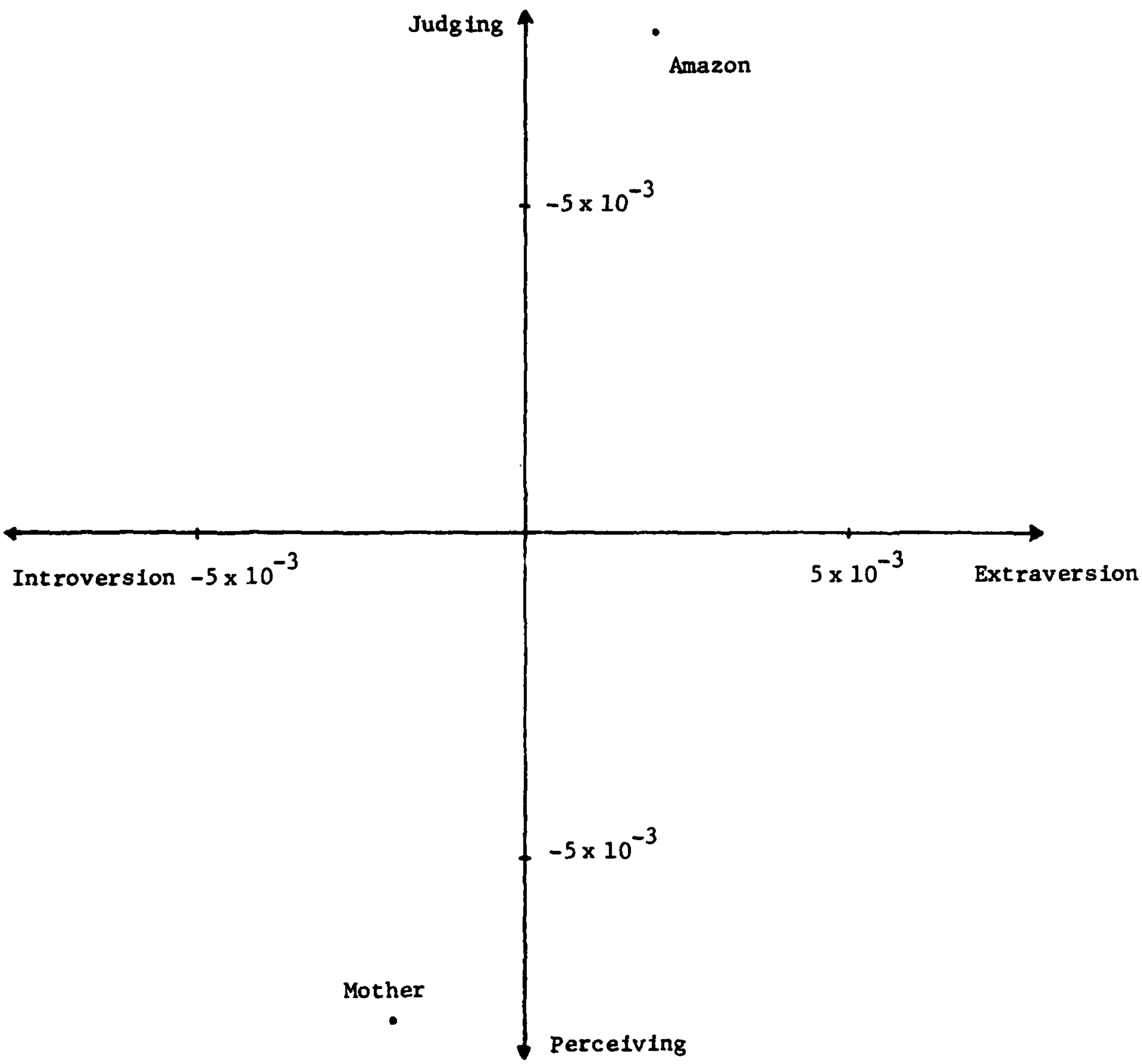

Figure 3. Location of the features of the item design matrix in the subject design space.

Table 2. The Fisher Information Matrix of the Coefficients.

		(Mother-Hetaira)			(Amazon-Medium)		
		x(Intercept)	x(E-I)	x(J-P)	x(Intercept)	x(E-I)	x(J-P)
M-H	x Intercept	.529					
	x E-I	.00454	.0314				
	x J-P	-.00432	.0196	.0306			
A-M	x Intercept	.448	.00349	-.00303	.528		
	x E-I	.00293	.0312	.0196	.00507	.0314	
	x J-P	-.00368	.0195	.0304	-.00364	.0195	.0304

VII. Conclusion.

In the previous six chapters various aspects of a statistical methodology have been described; now, in closing, an overview seems appropriate. While never made explicit, the models developed here parallel those of classical multivariate statistical analyses. Three similarities are the following:

(1) The multivariate normal distribution is the central object of study in, say, Anderson (1958). Similarly, the sampling function $p_j(\cdot)$, derived in chapter I and parametrized in chapter II, occupies a key position. Naturally, the consequences of assuming the form of $p_j(\cdot)$ need to be critically evaluated, as do the consequences of assuming multivariate normality. The axiomatic development of chapter I is presented to elucidate some of these issues.

(2) The parametrization of $p_j(\cdot)$ in the initial section of chapter II is rather analogous in form to the multivariate general linear hypothesis (Anderson, chapter 8 [1958]) of classical multivariate analysis. Indeed, the practical import of both models is to pose to the consumer of statistical analyses a relatively simple problem: the specification of relevant predictors. A distinction between the two is that the Q-sort model presented here poses a "double" specification problem. Not only must relevant predictors (W) be specified, so also must descriptions of the items (Q) be provided.

(3) Toward this end, the factor models of chapter V are presented. In the instance when both design matrices are being estimated, the result

resembles Hotelling's canonical correlation analysis (see Anderson, chapter 12 [1958]). When, however, one of the two matrices is fixed, the result is more analogous to principal components (Anderson, chapter 11 [1958]). In practice, both principal components and canonical correlations are used to aid in reducing data and specifying models; hopefully, so shall these factor models.

Aside from paralleling classical multivariate analysis, the present work, in particular chapter I, contributes in a minor way to the body of mathematical models that describe preference and selection behavior. The sampling function derived in chapter I is sufficiently similar to Luce's model that comparisons are meaningful while sufficiently different that these comparisons are interesting.

The sampling function of the Q-sorting model compares to that of Luce on the following points: (1) Both models express a notion of "independence of irrelevant alternatives", but (2) only for Luce's model is strong stochastic transitivity an immediate consequence. (3) Both models conceptualize the preference ordering activity as stochastic but only the Q-sort model is palindrome invariant. (4) Finally, Luce's model is a direct consequence of assuming a stability to a changing "context", i.e. a changing assortment of "irrelevant" alternatives. The Q-sorting model makes no such assumption. For these reasons, the Q-sorting model is a useful theoretical "foil" against which Luce's model may be better understood. And as with all foils, it would be of considerably less interest were such contrasts not possible.

To sum up, this work contributes in two ways. First and primarily, it develops a methodology for analyzing Q-sort data. Second and secondarily, it adds a new aspect to the theoretical investigation of preference behavior.

References

Andersen, E.B. "Asymptotic properties of conditional maximum-likelihood estimates." Journal of the Royal Statistical Society, Series B. 1970. 32: 383-301.

Anderson, T.W. An Introduction to Multivariate Statistical Analysis. New York: Wiley, 1958.

Block, J. The Q-sort method in personality assessment and psychiatric research Springfield, Ill: Thomas, 1961.

Burt, C. "The reciprocity principle." In Science, Psychology, and Communication: Essays Honoring William Stephenson. Brown, S.R. and Brenner, D.J. eds. New York: Teachers College Press, 1972; 39-56.

Burt, C., and Stephenson, W. "Alternative views on correlations between persons." Psychometrika 1939. 4: 269-281.

Carroll, J.B. "Approximating simple structure in factor analysis." Psychometrika 1953. 18: 23-38.

Cronbach, L.J., Gleser, G.C., Nanda, H., and Rajaratnam, N. The Dependability of Behavioral Measurements: theory of generalizability for scores and profiles. New York: Wiley, 1972.

Ferguson, G.A. "The concept of parsimony in factor analysis." Psychometrika 1954. 19: 281-290.

Godambe, V.P. "Sufficiency and ancillarity with a nuisance parameter." Biometrika 1980. 67: 155-162.

Jackson, D.M. and Bidwell, C.E. "A modification of Q-technique." Educational and Psychological Measurement 1959. 19: 221-232.

Kaiser, H.F. "The varimax method of factor analysis." Unpublished Ph.D. dissertation. 1956. University of California, Berkeley, California.

Kerlinger, F.M. "Q methodology in behavioral research." In Science, Psychology, and Communication: Essays Honoring William Stephenson Brown, S.R. and Brenner, D.J. eds. New York: Teachers College Press, 1972, 3-28.

Luce, R.D. Individual choice behavior. New York: Wiley, 1959.

Luce, R.D. "The choice axiom after twenty years." Journal of Mathematical Psychology 1977. 15: 215-233.

McCullagh, P. "A class of parametric models for the analysis of square contingency tables with ordered categories." Biometrika 1978. 65: 413-418.

Marley, A.A.J. "Some probabilistic models of simple choice and ranking." Journal of Mathematical Psychology 1968. 5: 311-332.

Neuhaus, J.O. and Wrigley, C. "The quartimax method: An analytical approach to orthogonal simple structure." British Journal of Mathematical and Statistical Psychology 1954. 7: 81-91.

Neyman, J. and Scott, E.L. "Consistent estimates based on partially consistent observations." Econometrika 1948. 16: 1-32.

Overall, J.E. and Klett, C.J. Applied Multivariate Analysis. San Fransicso: McGraw-Hill, 1972.

Saunders, D.R. "An analytic method for rotation to orthogonal simple structure." Research Bulletin 53-10. 1953. Princeton, N.J.: Educational Testing Service.

Sherlock, P.K. "The Relationship to Wolff's Archetypal Feminine Images to Time Orientation and Related Psychological Variables." Unpublished Ph.D. dissertation. 1980. Pacific Graduate School of Psychology, Palo Alto, California.

Spranger, E. Types of Men: The Psychology and Ethics of Personality. Halle (Saale), Germany. M. Neimeyer, 1928.

Stephenson, W. The Study of Behavior: Q-technique and its Methodology Chicago: University of Chicago Press, 1953.

Sunland, D.M. "The construction of Q-sorts: A criticism." Psychological Review 1962. 69: 62-64.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 303	2. GOVT ACCESSION NO. AD-A102 169	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A PARAMETRIC ANALYSIS OF STRUCTURED AND UNSTRUCTURED Q-SORT DATA		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) WILLIAM D. HEAVLIN		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0475
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office Of Naval Research Statistics & Probability Program Code 436 Arlington, VA 22217		12. REPORT DATE June 9, 1981
		13. NUMBER OF PAGES 69
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Q-sort behavior models; conditional likelihood, exponential families.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

PLEASE SEE REVERSE SIDE.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TR # 303

A PARAMETRIC ANALYSIS OF STRUCTURED AND UNSTRUCTURED Q-SORT DATA

The operation of sorting the items of a Q-set according to their similarity to a given object is idealized by a system of axioms. As a consequence of this axiom system, a stochastic model of Q-sorting behavior is derived. This model, with its associated axioms, resembles in some respects the preference model of Luce; the two are compared at length.

This Q-sort model is easily embedded in an exponential family. The domain of this family is the discrete space of all permutations of the items in the Q-set while its natural parameter space consists of the coefficients of a predictive linear relation. To operationalize the model, not only must predictor variables be specified, so also must a design matrix describing the underlying structure of the Q-set be furnished. Such a constuction occurs naturally for structured Q-sort data, but must be derived empirically for unstructured Q-sort data. Some empirical methods for deriving a description of the underlying structure of the Q-set are developed in the context of a "factor analysis" model.

The normalizing constant of this exponential family is impossible to compute; the calculation of a useful surrogate for this constant is the primary technical problem of this work. The solution posed considers a conditional likelihood whose normalizing constant is calculated over an intuitively appealing subgroup of permutations. A computer implementation of an algorithm for maximizing such a conditional likelihood model is described.

DAT
FILM